AD________________

Award Number: DAMD17-98-1-8499

TITLE: The Single Prostate Cell Transcriptome as Biological Assay

PRINCIPAL INVESTIGATOR: Peter S. Nelson, M.D.

CONTRACTING ORGANIZATION: Fred Hutchinson Cancer Research Center
Seattle, Washington 98109-1024

REPORT DATE: March 2001

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE
March 2001

3. REPORT TYPE AND DATES COVERED
Annual (1 Oct 99 - 28 Feb 01)

4. TITLE AND SUBTITLE
The Single Prostate Cell Transcriptome as Biological Assay

5. FUNDING NUMBERS
DAMD17-98-1-8499

6. AUTHOR(S)
Peter S. Nelson, M.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Fred Hutchinson Cancer Research Center
Seattle, Washington 98109-1024

E-MAIL:
pnelson@fhcrc.org

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

Technological advances in robotics and image analysis software have expanded the capabilities of large-scale strategies to provide for a more complete analysis of molecular differences between tissues or cells of interest. The research accomplished to date has demonstrated the ability to reproducibly isolate defined prostate cell populations by microdissection and flow cytometry. Gene expression studies of the cells purified by flow-cytometry reveal an altered expression profile that we believe results from the tissue dissociation/dispersion procedures. Ongoing and future work employs microdissection as the procedure of choice for specific cell-type analyses. The microdissection approach using a laser capture microscope is an efficient procedure for isolating cells representing abundant cell types, and we have isolated, purified, and analyzed the gene expression profiles from luminal epithelium and stromal elements. We have greatly expanded the database of sequences acquired from specific prostate cell types, and constructed arrays encompassing a wide range of diverse genes (n=6,000). In preliminary experiments we have used amplified cDNA probes isolated from small cell numbers to assess the gene expression profiles of defined cell types.

14. SUBJECT TERMS
Prostate Cancer, transcriptome, cDNA, microdissection

15. NUMBER OF PAGES
27

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

PN
X Where copyrighted material is quoted, permission has been obtained to use such material.

PN
X Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

PN
X Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

N/A In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

PN
X For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

N/A In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

N/A In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

04-23-01

PI - Signature Date

TABLE OF CONTENTS **PAGE**

INTRODUCTION

The subject of the research in this proposal is to develop methods for the examination of molecular alterations in prostate cancer at the level of single cells. The purpose of the research is to use these methods to identify molecular alterations in prostate cancer cells that can be used either singly or in combination to provide insights into the molecular evolution of prostate carcinogenesis, and produce a set of molecular tools capable of influencing the clinical management of patients with prostate carcinoma. The scope to the research involves the construction of cDNA libraries representing the genes expressed in selected populations of normal and neoplastic prostate cancer cells followed by the construction of microarrays suitable for comprehensive gene expression studies. These arrays will then be used to evaluate methods for single-cell transcriptome amplification with the aim of identifying a cohort of cellular transcripts which reflect a cellular phenotype.

BODY

Technical objective 1: To *obtain defined populations of normal and neoplastic prostate cell types which retain in-situ cellular characteristics*

- *Task 1: obtain and pathologically characterize fresh samples of normal, primary neoplastic, and metastatic carcinoma. Prepare tissue sections in frozen and fixed formats. Perform immunohistochemistry.* Completed.

- *Task 2: purify normal luminal, normal basal, and primary carcinoma cell populations using flow cytometric sorting. Disaggregate tissues, immuno-label, sort, assess sorted populations for purity via microscopic examination and by PCR analysis. Sort single cells into microtiter format.* We have sorted and purified normal basal and luminal cells by flow cytometry and constructed a cDNA library from each population (described in the previous report). We have sorted primary carcinoma cell populations (manuscript in preparation; Liu et al). Isolation of RNA from the purified cell populations has been inconsistent in terms of quality and quantity. The work is ongoing to optimize the methods using alternative RNA preservation reagents (e.g. RNAlater).

- *Task 3: evaluate alternative tissue digestion protocols.* We have disaggregated tissue samples with trypsin, with EDTA alone, and with Dispase without a significant improvement in quality/quantity of RNA extraction compared to the standard collagenase protocol. Gene expression alterations resulting from the dissagregation procedure remain a major hurdle for using this approach with flow cytometry as a means to profile gene expression from solid tissues.

- *Task 4: microdissect cohorts of phenotypically distinct prostate cells: luminal epithelium, basal epithelium, PIN, carcinoma foci, metastatic foci.* We have employed a new approach for microdissection that uses a laser-capture microscope (Arcturus) and used this methodology to constuct 3 new prostate cDNA libraries; one representing prostate basal cells; one representing prostate luminal cells; and one representing prostate stromal cells. Following the development of protocols aimed at optimizing both laser capture microdissection and RNA isolation, 24,000 cells each of stroma, luminal epithelium, and basal epithelium were captured and the RNA isolated by spin-column purification methods. cDNA libraries were constructed in a λ-phage vector using Clontech's SMART cDNA-PCR method. The respective libraries were then converted into phagemids and 300-700 clones from each library were sequenced for initial library characterization (Table 1). Genes specific to each cell type were identified including PSA from the luminal cell library, PSCA from the basal cell library, and vimentin from the stromal cell library. Furthermore, optimized microscopy methods, combining both laser *catapult* microdissection and laser *capture* microdissection, are now in place so as to improve capture of basal epithelial cells: cells more difficult to isolate than other prostate cell-types. To date, there are no published reports characterizing or comparing and contrasting the gene expression profiles and/or cDNA library construction from populations of prostate stromal, luminal epithelial, or basal epithelial cells. A manuscript detailing these libraries is in preparation (see reportable outcomes, Moore et al). We anticipate that these libraries will be useful tools for a variety of applications, including identifying prostate-specific genes, cell type-specific genes within the prostate, and in differential gene expression analysis. Library construction from PIN, primary carcinoma, and metastatic carcinoma are in progress.

Table 1. Sequence analysis of cDNA libraries constructed after cell isolation by Laser Capture Microscopy.

	Luminal Epithelium	Basal Epithelium	Stroma
No. clones sequenced	768	288	741
% w/o annotations	27	55	28
% annotated	73 (557)	45 (130)	72 (534)
% mitochondrial	18	34	11
% ribosomal	3	10	3.7

- *Task 5: microdissect single cells (20) from each of the above-described phenotypes.*
While we have been able to consistently *isolate* single cells from prostate cancer sections using laser capture microdissection, the ability to amplify the amount of cDNA needed for use in cDNA library construction or cDNA microarray analysis from the limited amount of RNA available in single cells remains challenging. We are continuing to investigate techniques, including amplification of mRNA that will allow us to develop comparative gene expression profiles from individual prostate cells. At the same time, we have made progress in developing the tools necessary for the analyses of molecular changes within individual prostate epithelial cells isolated from peripheral blood, apheresis samples, and/or bone marrow. These tools include immunostaining cell preparations (epithelial cells isolated from periph-

eral blood, apheresis samples, or bone marrow using positive and negative selection methods) for prostate specific antigen, capturing positively-stained cells by laser capture microdissection, cell lysis, and DNA analysis for methylation of GST-pi and androgen receptor mutations. We have successfully amplified and sequenced exon 8 of the androgen receptor from individual circulating prostate cells and have identified no mutations in a cohort of 5 patients. We are currently developing a protocol to determine GST-pi methylation status.

- *Task 6: assess RNA quality (preservation) between frozen sections and fixed/stained sections.* As anticipated, our work in this area has demonstrated that the yield of RNA from frozen tissues is much greater and of higher quality than from comparable quantities of formalin fixed tissue. Our current protocol employs a rapid ethanol fixation of frozen tissue with or without an H&E or immunostain prior to LCM. We have successfully isolated intact RNA from formalin-fixed tissues, but to date this remains poorly reproducible.

- *Task 7: assess feasibility of flow sorted single cell isolation automation.* We are not currently pursuing this approach due to the alterations in gene expression resulting from tissue disagreggation. Future work may entail flow cytometric isolation of epithelial cells in peripheral blood or bone marrow.

- *Task 8: (future work) refine cell phenotype acquisition based upon the development of new markers/antibodies.* In collaboration with Dr. Alvin Liu in the Department of Molecular Biotechnology, we have identified several additional antigens recognized with monoclonal antibodies that can be used for sorting prostate epithelial cells by flow cytometry (see reportable outcomes, Liu et al). The future application of these discriminating proteins/antigens will await the development of consistent amplification protocols as described in this proposal.

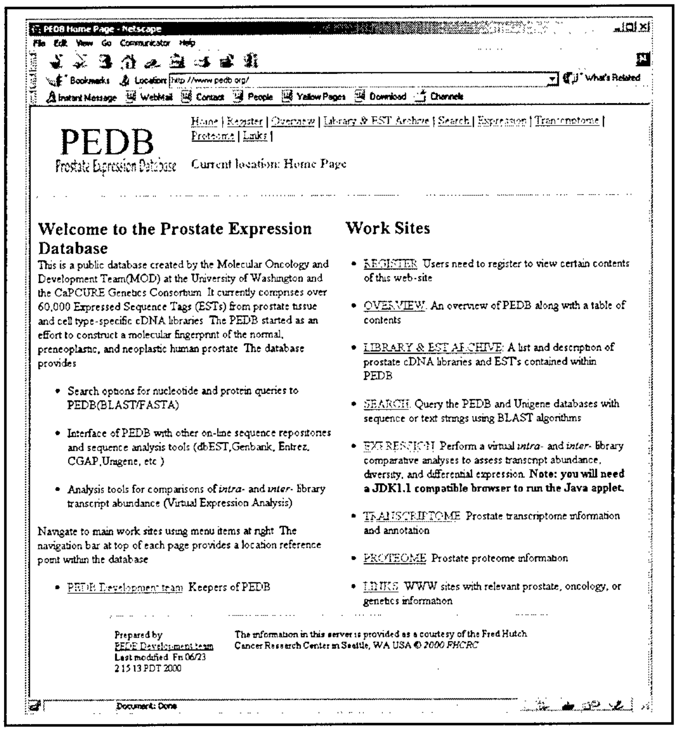
Technical objective 2: *To construct microarrays of prostate transcripts that reflect the gene expression potential of the cell types to be examined.*

- *Task 8: identify a non-redundant clone set from the Prostate Expression Database to encompass all highly expressed transcripts (~12), moderately expressed transcripts (~500) and several thousand rare transcripts (~6000).*
  We have now identified and assembled a non-redundant set of 6,000 cDNAs (ESTs) from the prostate expression database that are suitable for array construction. Many of these genes are derived from the cell type-specific libraries described above.

- *Task 9: retrieve cDNA clones from archive, PCR amplify inserts with amine-linked primer, and purify.* We have retrieved 6,000 cDNA clones from the cDNA archive and amplified the inserts by PCR. Our current array construction methodology at the Fred Hutchinson Cancer Center uses poly-lysine coated slides and demonstrates excellent reproducibility and sensitivity. We have used these arrays for the identification of genes in the prostate under the control of the androgen receptor and androgenic ligands (See reportable outcomes, Lin et al).

- *Task 10: construct 3 normalized cDNA libraries from flow sorted basal, luminal, and primary carcinoma (CD44+) without amplification procedures, and evaluate libraries for quality: diversity and abundance of transcripts.*
  As described in the previous report, we have constructed cDNA libraries from flow sorted basal (CD44+), luminal (CD57+), and primary carcinoma (CD44+) cells. We have also now constructed cDNA libraries from normal basal and luminal epithelial cells using microdissection approaches. A total of 2,500 ESTs have been produced from these libraries and entered into the Prostate Expression Database (www.pedb.org).

- *Task 11: pick random cDNA clones from the new libraries, array on nylon membranes and screen for abundant prostate cDNAs, select non-abundant species, PCR amplify inserts.* Random sequencing of cDNA clones from the libraries described above and additional libraries from prostate cancer cell lines has identified >18,000 distinct genes expressed in prostate tissues. We have used a virtual selection approach rather than the physical negative selection approach, to identify non-redundant clone sets representing the prostate transcriptome. These clones have been extracted from the database archive, re-arrayed into 384-well microtiter plates, amplified by PCR, and spotted onto microscope slides for subsequent hybridization.



Home | Register | Overview | Library & EST Archive | Search | Expression | Transcriptome | Proteome | Links |

PEDB
Prostate Expression Database

Current location: Home Page

Welcome to the Prostate Expression Database

This is a public database created by the Molecular Oncology and Development Team(MOD) at the University of Washington and the CaPCURE Genetics Consortium It currently comprises over 60,000 Expressed Sequence Tags (ESTs) from prostate tissue and cell type-specific cDNA libraries The PEDB started as an effort to construct a molecular fingerprint of the normal, preneoplastic, and neoplastic human prostate The database provides

- Search options for nucleotide and protein queries to PEDB(BLAST/FASTA)
- Interface of PEDB with other on-line sequence repositories and sequence analysis tools (dbEST,Genbank, Entrez, CGAP,Unigene, etc)
- Analysis tools for comparisons of *intra-* and *inter-* library transcript abundance (Virtual Expression Analysis)

Navigate to main work sites using menu items at right The navigation bar at top of each page provides a location reference point within the database

- PEDB Development team Keepers of PEDB

Work Sites

- REGISTER Users need to register to view certain contents of this web-site
- OVERVIEW. An overview of PEDB along with a table of contents
- LIBRARY & EST ARCHIVE: A list and description of prostate cDNA libraries and EST's contained within PEDB
- SEARCH. Query the PEDB and Unigene databases with sequence or text strings using BLAST algorithms
- EXPRESSION Perform a virtual *intra-* and *inter-* library comparative analyses to assess transcript abundance, diversity, and differential expression. **Note: you will need a JDK1.1 compatible browser to run the Java applet.**
- TRANSCRIPTOME Prostate transcriptome information and annotation
- PROTEOME Prostate proteome information
- LINKS WWW sites with relevant prostate, oncology, or genetics information

Prepared by
PEDB Development team
Last modified Fri 06/23
2 15 13 PDT 2000

The information in this server is provided as a courtesy of the Fred Hutch Cancer Research Center in Seattle, WA USA © *2000 FHCRC*

Figure 1. PEDB Web Interface. ESTs derived from cDNA clones sequenced from the microdissected prostate cell types are submitted to the Prostate Expression Database (PEDB) for processing. The sequences are assembled and annotated. A non-redundant set of 6,000 ESTs representing 6,000 different genes expressed in prostate tissues was identified. The cDNA clones were re-distributed in microtiter plates, amplified, and spotted onto glass microscope slides for microarray hybridization experiments.

- *Task 12: construct physical micro-arrays of cDNA clones on glass supports using robotic tools: total of 500 replicates.* See Task 9 above. We are currently using a GeneMachines robotic spotting tool with the capability of spotting >18,000 cDNAs per microscope slide. More than 500 replicate slides have been printed to date comprising the 6,000 prostate PEDB cDNAs. The current use for these slides is for the analysis of amplification procedures in order to assess the fidelity of probe material obtained from small numbers of cells.

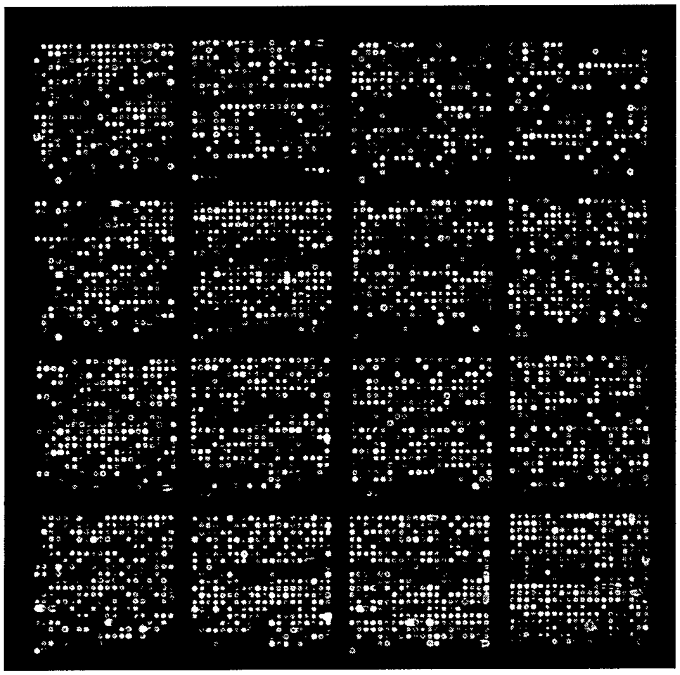


Figure 2. Microarray hybridization with amplified cDNA. A portion of the 6,000-clone PEDB cDNA microarray hybridized with amplified cDNA from the LNCaP prostate cancer cell line. A reference standard of pooled RNA from 3 different prostate cancer cell lines serves as the control. Red spots indicated up-regulated genes and green spots indicate down-regulated genes in the experiment relative to control.

- *Task 13: assess alternative array methodologies as they become available (ink jet oligonucleotide)* This task will be on-going for the duration of the proposal.

<u>**Technical objective 3:**</u> *To construct representative probes from single or small numbers of defined cells that are suitable for micro-array interrogation, and retain the transcriptome composition (diversity and abundance) present in the original cell type(s).*

- Task 14: convert to cDNA, amplify by PCR, and label nucleic acid from flow sorted cell populations of decreasing cell quantities. Assess quality by Northern analysis and hybridization to small "known clone" array. Compare with unamplified "traditional" probe. (months 12-13). We are not currently using the flow-cytometry isolation approach due to changes in gene expression associated with tissue disruption. Our focus now centers on microdissection.

- Task 15: as above with microdissected populations. (months 13-14). We have successfully microdissected prostate luminal and basal cells from 10 μm frozen sections. Amplification using the PCR-based strategy incorporating an anchored primer has been successful in producing adequate amounts of cDNA for probe construction and hybridization. However, the fidelity of the amplification in our first attempts has been poor. This results in the skewing of message abundance levels in the probe material relative to the starting material. Our next approach will be to truncate the cDNAs roughly to a common size using a frequent cutting restriction enzyme followed by adapter ligation and subsequent amplification. This approach is designed to eliminate length bias that may be skewing message ratios.
- Task 16: as above with aRNA method and flow sorted cells (months 15-16). As described above, we are not currently using flow-sorting for cell isolation and probe construction.
- Task 17: as above with microdissected populations. (months 17-18). We have used a modification of the aRNA protocol developed by Eberwine et al. We have achieved a ~100-fold amplification with a first round aRNA synthesis and an additional ~100-fold amplification with a second round. This allows for the use of ~0.5 ng of total RNA for probe construction. However, the aRNA amplification is still not suitable for the analysis of single, or small numbers of microdissected cells. A second experiment using shorter RNA polymerase-mediated synthesis duration looks promising. These data are currently undergoing analysis.
- Task 18: as above with microdissected populations from frozen and fixed tissues. (months 19-20). Work in progress.
- Task 19: convert to cDNA, amplify, label, and hybridize single-cell probes to high-density oligonucleotide arrays. (months 21-25). Work pending.
- Task 20: capture and quantitate hybridization spot intensities on fluorimage laser scanners, and enter into database. (months 21-25). We have contacted investigators at Stanford University and acquired software for the incorporation of cDNA array data into PEDB. The software (database architecture) is currently being incorporated into the structure of PEDB and should facilitate the storage and analysis of microarray data.

Technical objective 4: *To identify a cohort of cellular transcripts which correlate with, define, or "fingerprint", a cellular phenotype(s).*

The following work is in progress or pending the completion of Technical objective 3.

- Task 21: examine hybridization intensities (values) for each datapoint in an automated, comparative fashion from cells of *a priori* defined identical phenotype (luminal epithelium with luminal epithelium) to develop cohorts of phenotype-defining transcripts. (months 26-27)
- Task 22: examine hybridization intensities between cells with *a priori* defined different phenotypes to establish a lineage relationship. (months 26-27)
- Task 23: correlate expression profiles with known molecular/biochemical/functional data concerning each cell type (metastatic location). (months 27-28)
- Task 24: analyze by DNA sequencing cDNAs which are in phenotype cohorts and have not previously been defined. (months 26-28)
- Task 25: analyze expression data using cluster and phylogeny algorithms to assess lineage relationships. (months 26-29)
- Task 26: plan molecular experiments and clinical evaluation of candidate phenotype-defining cohorts: e.g.1) retrospective analysis of carcinomas with known clinical outcomes (progres-

sion/metastasis) 2) prospective analysis diagnostic needle biopsy samples 3) evaluation of unrecognized or "latent" cancer samples obtained at autopsy. (months 27-30)

- Task 26: analyze/compile data and prepare formal report (month 30)

KEY RESEARCH ACCOMPLISHMENTS (since the previous annual report)

- Obtained and purified single circulating neoplastic prostate cells from the peripheral blood of patients with prostate cancer and analyzed exon 8 of the androgen receptor for molecular alterations.
- Constructed cDNA libraries from laser-capture microdissected prostate luminal and basal epithelial cells and prostate stroma.
- Sequenced and analyzed 2,000 cDNAs (producing ESTs) from the luminal cell, basal cell, and stromal libraries.
- Constructed cDNA microarrays comprised of 6,000 different prostate cDNAs.
- Constructed complex cDNA probes from microdissected cells and 0.5ng total RNA and used the probe in microarray hybridization experiments. The amplification fidelity is yet to retain the appropriate gene expression ratios suitable for experimental comparisons.
- Acquired database software for archiving and analyzing cDNA microarray experiments.

REPORTABLE OUTCOMES

Nelson PS, Han D, Rochon Y, Corthals G, Lin B, Monson A, Nguyen V, Franza BR, Plymate SR, Aebersold R, and Hood L. (2000) Comprehensive analyses of prostate gene expression: convergence of EST databases, transcript profiling and proteomics. *Electrophoresis* 21:1823-31.

Grouse LH, Munson PJ, and **Nelson PS**. (2001) Sequence Databases and Microarrays as Tools for Identifying Prostate Cancer Biomarkers. *J. Urology* 57 (Suppl 4A): 154-159.

Li P and **Nelson PS**. (2001) Prostate Cancer Genomics. *Current Urology.*

Clegg N, Erolgu B, Ferguson C, Arnold H, Mooreman A, and **Nelson PS**. Digital Expression Profiles of the Prostate Cell Transcriptome (Submitted: *Genomics*)

Liu AY, **Nelson PS**, Ligner CL, van den Engh G, Hood L. Human Prostate Epithelial Cell-Type cDNA Libraries and Expression Pattern in Prostate Cancer. (Submitted: *Cancer Research*)

Lin B, White JT, Arnold H, Ferguson C, Hood L, and **Nelson PS**. The Program of Androgen-Responsive Genes in the Human Prostate (In preparation).

Moore S and **Nelson PS.** Comparative Analysis of the Luminal and Basal Epithelial Cell Transcriptomes from the Normal and Neoplastic Prostate. (In preparation).

CONCLUSIONS

The research accomplished to date has demonstrated the ability to reproducibly isolate defined prostate cell populations by microdissection and flow cytometry. Gene expression studies of the cells purified by flow-cytometry reveal an altered expression profile that we believe results from the tissue dissociation/dispersion procedures. Ongoing and future work employs microdissection as the procedure of choice for specific cell-type analyses. The microdissection approach using a laser capture microscope is an efficient procedure for isolating cells representing abundant cell types, and we have isolated, purified, and analyzed the gene expression profiles from luminal epithelium and stromal elements. We have greatly expanded the database of sequences acquired from specific prostate cell types, and constructed arrays encompassing a wide range of diverse genes (n=6,000). In preliminary experiments we have used amplified cDNA probes isolated from small cell numbers to assess the gene expression profiles of defined cell types.

REFERENCES

None

APPENDICES

Nelson PS, Han D, Rochon Y, Corthals G, Lin B, Monson A, Nguyen V, Franza BR, Plymate SR, Aebersold R, and Hood L. (2000) Comprehensive analyses of prostate gene expression: convergence of EST databases, transcript profiling and proteomics.*Electrophoresis* 21:1823-31.

Grouse LH, Munson PJ, and **Nelson PS**. (2001) Sequence Databases and Microarrays as Tools for Identifying Prostate Cancer Biomarkers. *J. Urology* 57 (Suppl 4A): 154-159.

Peter S. Nelson[1,2,3]
David Han[2]
Yvan Rochon[2]
Garry L. Corthals[4]
Biaoyang Lin[2]
Adam Monson[2]
Vilaska Nguyen[2]
B. Robert Franza[2,5]
Stephen R. Plymate[3]
Ruedi Aebersold[2]
Leroy Hood[2]

[1]Division of Human Biology,
Fred Hutchinson Cancer
Research Center,
Seattle, WA, USA
[2]Department of Molecular
Biotechnology, University
of Washington,
Seattle, WA, USA
[3]Department of Medicine,
University of Washington,
Seattle, WA, USA
[4]Garvan Institute for Medical
Research, Sydney, NSW,
Australia
[5]Department of Cell
Systems Initiative,
University of Washington,
Seattle, WA, USA

Comprehensive analyses of prostate gene expression: Convergence of expressed sequence tag databases, transcript profiling and proteomics

Several methods have been developed for the comprehensive analysis of gene expression in complex biological systems. Generally these procedures assess either a portion of the cellular transcriptome or a portion of the cellular proteome. Each approach has distinct conceptual and methodological advantages and disadvantages. We have investigated the application of both methods to characterize the gene expression pathway mediated by androgens and the androgen receptor in prostate cancer cells. This pathway is of critical importance for the development and progression of prostate cancer. Of clinical importance, modulation of androgens remains the mainstay of treatment for patients with advanced disease. To facilitate global gene expression studies we have first sought to define the prostate transcriptome by assembling and annotating prostate-derived expressed sequence tags (ESTs). A total of 55 000 prostate ESTs were assembled into a set of 15 953 clusters putatively representing 15 953 distinct transcripts. These clusters were used to construct cDNA microarrays suitable for examining the androgen-response pathway at the level of transcription. The expression of 20 genes was found to be induced by androgens. This cohort included known androgen-regulated genes such as prostate-specific antigen (PSA) and several novel complementary DNAs (cDNAs). Protein expression profiles of androgen-stimulated prostate cancer cells were generated by two-dimensional electrophoresis (2-DE). Mass spectrometric analysis of androgen-regulated proteins in these cells identified the metastasis-suppressor gene NDKA/nm23, a finding that may explain a marked reduction in metastatic potential when these cells express a functional androgen receptor pathway.

Keywords: Prostate / Transcriptome / Proteome / Androgen / Microarray EL 3957

1 Introduction

The development and subsequent progression of human prostate carcinoma is propelled by the accumulation of genetic alterations and influenced by environmental factors. One pivotal mediator of prostate carcinogenesis is the androgen receptor (AR) pathway. The majority of prostate cancers initially require androgens for growth, and the elimination of AR-ligands by surgical or chemical castration leads to marked tumor regression through a mechanism of programmed cell death [1]. The manipulation of the AR pathway has been used in clinical medicine since the 1940s as the primary treatment of advanced prostate cancer. However, this therapy is palliative and eliminates the potential beneficial effects of androgen-induced cellular differentiation. Surviving cancer cells lose their dependence on androgens over time and are capable of prolifertion in the absence of serum androgens. The molecular events leading to androgen independence (AI) have not been defined, but potential mechanisms include overexpression of the AR, mutations in the AR gene leading to promiscuous ligand binding, and the activation of the AR or downstream regulatory molecules by other endocrine or paracrine growth factors [2, 3].

Until recently, biological investigations have almost entirely focused on the study of individual genes and proteins. This has partly been due to the submicroscopic nature and transient existence of relevant molecules, combined with a lack of quantitative technology capable of providing accurate comprehensive views of biological complexity. Significant advances have been achieved studying individual genes, proteins and small numbers of molecular interactions. However, conclusions made on the basis of the study of an individual gene may have limited relevance as to how the gene and gene product function in the context of the whole cell, tissue, or organism. Progress in understanding complex molecular processes

Correspondence: Peter Nelson, M.D., Division of Human Biology, Mailstop D4-100, 1100 Fairview Avenue, Seattle, WA 98109-1024, USA
E-mail: pnelson@fhcrc.org
Fax: +206-667-2917

Abbreviations: AR, androgen receptor; **cDNA**, complementary DNA; **μLC**, microcapillary LC; **PEDB**, prostate expression database; **PSA**, prostate specific antigen

0173-0835/00/0909-1823 $17.50+.50/0

has been hampered by the lack of a complete inventory or "tool-set" of all genes and their cognate proteins that are necessary for defining phenotypes of normal and pathological cellular states.

The completion of the Human Genome Project will provide a foundation for a thorough description of this molecular inventory. More specifically, the gene inventory or tool set required for studies of prostate carcinogenesis is that portion of the human genome used or expressed in the human prostate gland. The subset of genes transcribed or expressed in a given cell or tissue type such as the prostate may be defined as the "transcriptome", the dynamic link between the genome, the proteome, and the cellular phenotype associated with physical characteristics [4]. Once a transcriptome has been described, the next objective is to understand the relationships of the genes and their protein products in terms of a complex system, *e.g.*, biological pathways and networks, that may define health and disease. With this aim, novel technologies for comprehensively assessing genomes and patterns of gene expression have recently been developed.

Our initial efforts have focused on defining the prostate transcriptome through the production and assembly of expressed sequence tags (ESTs) derived from prostate complementary DNA (cDNA) libraries representing a wide sprectrum of normal and neoplastic states. These EST assemblies have been used to construct cDNA microarrays suitable for interrogating the transcriptome in experiments designed to examine specific biological pathways that may be involved in prostate carcinogenesis. The molecular pathway mediating androgenic hormone action on prostate cells is a specific focus of our work. The functional architecture of prostate gene networks is furth elucidated by our next level of analysis that incorporates studies of the prostate proteome. Analysis of the transcriptome facilitates proteome studies by providing a comprehensive prostate sequence database for identifying and annotating known and unknown proteins displayed by two-dimensional gel electrophoresis (2-DE) and analyzed by mass spectrometry (MS). Our objectives for delineating the molecular network(s) influenced by AR activation are to identify specific targets that promote the differentiation and apoptotic potential of prostate cancer cells while reducing their ability to proliferate.

2 Materials and methods

2.1 Assembly of a prostate transcriptome: Prostate Expression Database (PEDB)

A prostate transcriptome was assembled from ESTs derived from cDNA libraries representing a wide sprectrum of normal, benign, and malignant prostate tissues. A detailed description of the assembly and annotation procedure is described elsewhere [5]. Briefly, individual ESTs, detailed cDNA library information, and sequence annotations were loaded into a relational database (Oracle Corp.) termed the Prostate Expression Database (PEDB). Prostate ESTs used for the assembly were generated in our laboratory as previously described [6]. Additional public domain ESTs of prostate origin were obtained from Genbank (http://www.ncbi.nlm.nih.gov/Entrez/batch.html), the NCI Cancer Genome Anatomy Project (CGAP) [7], and The Institute for Genome Research (TIGR) (http://www.tigr.org). Each EST was examined for sequence homology to cloning vectors, *Escherichia coli*, and repetitive DNA sequences using a core program called AnalDemon (http://www.mbt.washington.edu/PEDB/software). AnalDemon first employs Cross_Match (http://bozeman.mbt.washington.edu/phrap.docs/general.html); a program based on the Smith-Waterman-Gotoh algorithm, to screen for vector and *E. coli* genome contamination. ESTs are then examined for interspersed repeats and regions of low sequence complexity using Repeatmasker (http://ftp.genome.washington.edu/RM/RepeatMasker.html). Specific portions of EST sequences exhibiting homology to any of these unwanted elements are masked in order to eliminate the sequence from contributing to an assembly process. CAP2 [8], a multiple alignment program based on a variant of the Smith and Waterman algorithm, was used for assembling ESTs into homologous groups or clusters. Clustering is based on maximal scoring of overlapping alignments and allows for general substitutions resulting from sequencing errors, insertions, and deletions. CAP2 produces a consensus sequence and allows varying sensitivity and overlap parameters. Each group or cluster of ESTs exhibiting significant homology with one another is termed a species. Thus, a species is a sequence or group of sequences that is unique relative to the nucleotide sequence of other groups of sequences, and each is given a unique PEDB Species Identification number (SID). The SID provides a means to analyze gene expression across the entire set of assemblies, and can be used to provide a library-by-library species-specific differential expression profile. Each distinct species from the clustering process was annotated by searching the Unigene (ncbi.nlm.nih.gov in / pub/schuler/unigene), Genbank (ncbi.nlm.nih.gov/blast/db/nt.Z), and EST databases (ncbi.nlm.nih.gov/blast/db/est.Z) using BLASTN (http://blast.wustl.edu). Annotations were assigned automatically using the program Smart-Blast (http://www.mbt.washington.edu/PEDB/software) by selecting the database match with the lowest *p* value and the highest blast score where the maximum *p* value is e^{-20} and the minimum blast score is 500.

2.2 Prostate transcriptome analyses by cDNA microarray

2.2.1 Microarray fabrication

A nonredundant set of 1500 prostate-derived cDNA clones was identified from the prostate transcriptome archived in PEDB. Individual clone inserts were amplified by the PCR using 2 µL of bacterial transformant culture as template with primers BL_m13F (5′-GTAAAACGA-CGGCCAGTGAATTG-3′) and BL_m13R (5′-ACACAGG-AAACAGCTATGACCATG-3′ as previously described [6]. PCR products were purified through Sephacryl S500 (Amersham Pharmacia Biotech, Uppsala, Sweden), mixed 1:1 with dimethylsulfoxide, and spotted in duplicate onto coated Type VII glass microscope slides (Amersham Pharmacia Biotech) using a Molecular Dynamics (Sunnyvale, CA, USA) GenII robotic spotting tool. After spotting, the glass slides were air-dried and UV-cross-linked with 500 mJ of energy and then baked at 95°C for 30 min.

2.2.2 Probe construction and microarray hybridization

Total RNA was isolated from the androgen-responsive LNCaP prostate cancer cells [9] at time points of 0, 4, 8, 24, and 72 h after androgen depletion or supplementation using TRIzol (Life Technologies, Paisley, UK) according to the manufacturer's directions. Fluorescence-labeled probes were made from 30 µg of total RNA in a reaction volume of 20 µL containing 1 µL anchored oligo-dT primer (Amersham Pharmacia Biotech), 0.05 mM Cy3-dCTP (Amersham Pharmacia Biotech), 0.05 mM dCTP, 0.1 mM each dGTP, dATP, dTTP, and 200 U Superscript II reverse transcriptase (Life Technologies). Reactants were incubated at 42°C for 120 min followed by heating to 94°C for 3 min. Unlabeled RNA was hydrolyzed by the addition of 1 µL of 5 N NaOH and heating to 37°C for 10 min. One µL of 5 M HCl and 5 µL of 1 M Tris-HCl, pH 7.5, were added to neutralize the base. Unincorporated nucleotides and salts were removed by chromatography (Qiagen, Chatsworth, CA, USA), and the cDNA was eluted in 30 µL dH2O. One µg of dA/dT 12–18 (Amersham Pharmacia Biotech) and 1 µg of human Cot1 DNA (Life Technologies) were added to the probe, heat-denatured at 94°C for 5 min, combined with an equal volume of 2 × microarray hybridization solution (Amersham Pharmacia Biotech) and prehybridized at 50°C for 1 h. The mixture was then placed onto a microarray slide with a coverslip and hybridized in a humid chamber at 52°C for 16 h. The slides were washed once with 1 × sodium chloride and sodium citrate (SSC), 0.2% SDS at room temperature for 5 min and then twice with 0.1 × SSC, 0.2% SDS at room temperature for 10 min. After washing, the slide was rinsed in distilled water to remove trace salts and dried.

2.2.3 Image acquisition and data analyses

Fluorescence intensities of the immobilized targets were measured using a laser confocal microscope (Molecular Dynamics). Intensity data were integrated at a pixel resolution of 10 µm using approximately 20 pixels per spot, and recorded at 16 bits. Quantitative data were obtained with the SpotFinder Version 2.4 program written at the University of Washington. Local background hybridization signals were subtracted prior to comparing spot intensities and dtermining expression ratios. For each experiment, each cDNA was represented twice on each slide, and the experiments were performed in duplicate producing four data points per cDNA clone and hybridization probe. Intensity ratios for each cDNA clone hybridized with probes derived from androgen-stimulated LNCaP and androgen-starved LNCaP cells were calculated (stimulated intensity/starved intensity). Gene expression levels were considered significantly different between the two conditions if all four replicate spots for a given cDNA demonstrated a ratio > 2 or < 0.5, and the signal intensity was greater than two standard deviations above the image background. We have previously determined that expression ratios less than 1.5 are not reproducible in our system (datas not shown).

2.3 Prostate proteome analyses by 2-DE and MS

2.3.1 2-DE

LNCaP prostate cancer cells were grown under conditions of androgen stimulation or androgen starvation as described above. M12AR cells, a highly metastatic prostate cancer cell line derived from the serial passaging of SV40 immortalized prostate epithelial cells [10] and transfected with the AR were grown in serum-free DMEM high-glucose media (Life Technologies) supplemented with insulin, transferrin, selenium, and dexamethasone as previously described [11]. Cells were allowed to reach 80% confluency and then treated for 24 h with the same media supplemented with 10 nM R1881. Cells were washed once with PBS, scraped from plates with a rubber policeman and pelleted by centrifugation. Protein was harvested as described by Garrels and Franza [12]. Briefly, cell pellets were lysed in a buffer containing 0.3% SDS, 1% β-mercaptoethanol, and 50 mM Tris-HCl, pH 8.0, 100 µg/mL DNAase I, 50 µg/mL RNAase A, 5 mM $MgCl_2$, and heated for 1 min at 100°C. Harvested protein was lyophilized, resuspended in isoelectric focusing (IEF) gel rehydration solution, and stored at –80°C. Soluble proteins were run in the first dimension by using a commercial flatbed electrophoresis system (Multiphor II; Amersham Pharmacia Biotech). Nonlinear immobilized pH gradient (IPG) dry strips ranging from 3.0 to 10.0 (Amersham Pharmacia Biotech) were used for the first-dimensional separation. Forty micrograms of protein from whole-cell lysates were mixed with IPG strip rehydration buffer (8 M urea, 2% Nonidet P-40, 10 mM dithiothreitol), and 250–380 μL of solution (13 and 18 cm IPGs, respectively) was added to individual lanes of an IPG strip rehydration tray (Amersham Pharmacia Biotech). The strips were rehydrated at room temperature for 1 h. The samples were run at 300 V, 10 mA, 5 W for 2 h, ramped to 3500 V, 10 mA, 5 W over a period of 3 h, and then kept at 3500 V, 10 mA, 5 W for 15–19 h. Following IEF (60–70 kVh), the IPG strips were first reequilibrated for 8 min in a solution of 2% w/v dithiothreitol, 2% w/v SDS, 6 M urea, 30% w/v glycerol, 0.05 M Tris-HCl (pH 6.8) and subsequently for 4 min in a solution of 2.5% w/v iodoacetamide, 2% w/v SDS, 6 M urea, 30% w/v glycerol, 0.05 M Tris-HCl (pH 6.8) with a trace of bromophenol blue added for color. Following reequilibration, the strips were transferred and apposed to 10% polyacrylamide second-dimensional gels. Polyacrylamide gels were poured in casting stand with 10% acrylamide-2.67% piperazine diacrylamide-0.375 M Tris, pH 8.8, 0.1% w/v SDS, 0.05% w/v ammonium persulfate, 0.05% TEMED (*N,N,N′,N′*-tetramethylethylenediamine) in Milli-Q water (Millipore, Bedford, MA, USA). Second-dimensional gels (0.1 × 20 × 20 cm) were run in an apparatus supplied by Oxford Glycosciences (Abington, UK). Once the IPG strips were apposed to the second-dimensional gels, they were immediately run at a constant current of 50 mA at 500 V and 85 W for 20 min, followed by a constant current of 200 mA at 500 V and 85 W until the buffer front was 10–15 mm from the bottom of the gel. Gels were removed and silver stained according to the procedure of Blum *et al.* [13].

2.3.2 Protein identification by tandem mass spectrometry

Protein spots from gels were identified by tandem mass spectrometry (MS/MS) as previously described [14]. Spots from silver-stained gels were excised and in-gel tryptic peptides were separated by microcapillary LC (μLC) coupled to a tandem mass spectrometer (TSQ 7000; Finnigan, San Jose, CA). Peptide fragmentation spectra were generated in a data-dependent fashion. Spectra were searched against the composite OWL protein sequence database by using the computer program SEQUEST [15] and against the PEDB. A protein match was determined by comparing the number of peptides identified and their respective cross-correlation scores. Protein identifications were verified by comparison with theoretical molecular weights and isoelectric points.

3 Results and discussion

3.1 Prostate gene expression analyses: EST assemblies and annotation

ESTs produced from cDNA libraries derived from normal and neoplastic human prostate tissue samples were entered into the PEDB, an Oracle relational database running on a Sun SPARC workstation. The most recent PEDB build was assembled starting with 55 000 prostate ESTs produced from 42 cDNA libraries. Portions of EST sequences with homology to cloning vector, *E. coli* genomic DNA, and human repetitive DNA sequences

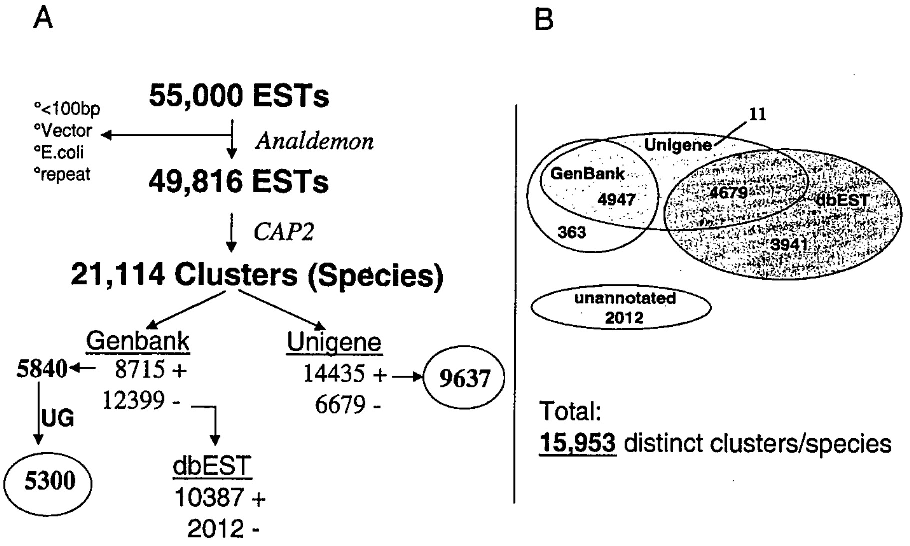


Figure 1. Assembly of a prostate transcriptome. (A) 55 000 prostate ESTs were examined for "junk" sequences leaving 49 816 high quality ESTs suitable for assembly. Clustering the ESTs into groups of high homology produced a set of 21 114 clusters that were annotated against nucleotide and protein sequences in the public sequence databases. Clusters exhibiting homology to Genbank sequences were also examined for homology to Unigene sequences (UG) to further collapse clusters into homologous groups. (B) Following clustering, database annotations and reclustering, a total of 15 953 distinct prostate EST species were identified. More than 2000 prostate species did not have homology to nonprostate-derived sequences in the public databases (unannotated).

were masked and ESTs with > 100 bp of high quality sequence were admitted to the assembly process (Fig. 1A). A total of 49 816 high quality ESTs were assembled using the sequence assembly program CAP2 to produce 21 114 clusters. Each cluster was annotated by searching the Unigene, Genbank, and dbEST databases with the CAP2-generated cluster consensus sequences using BLASTN. Clusters annotating to the same database sequence were joined to further reduce the number of distinct clusters to 15 953 (Fig. 1B).

Studies in the 1970s using reassociation kinetics to estimate the number of different transcripts indicate that between 10 000 and 30 000 distinct mRNAs are present in mammalian cells or organs [16, 17]. Recent data produced using the method of Serial Analysis of Gene Expression (SAGE) suport these estimates of transcript diversity in mammalian epithelial cells with estimates of 14 000–20 000 different mRNAs per cell [18]. Although the identification of alternatively spliced transcripts and highly homolgous gene family members may increase or decrease these estimates slightly, they nevertheless provide a rough estimate of the complexity of cellular gene activity. Based upon these data, the 15 953 prostate EST clusters that we have assembled should characterize roughly 50–75% of the prostate transcriptome. It is likely that this assembled dataset comprises all of the abundant and most of the moderately abundant prostate transcripts [6]. Ongoing work involves the acquisition of the remaining low abundance transcripts. Approaches to achieving this goal involve the construction of cDNA libraries from highly selected purified cell populations such as luminal epithelial and neuroendocrine cells, and from prostate tissues at different stages of development (*e.g.*, fetal prostate) or under different hormonal influences (*e.g.*, androgen stimulation). Another useful strategy involves the iterative removal of abundant and previously identified cDNAs in order to select for rare species. A high-throughput method using cDNA array-based technology has been developed to facilitate this process [19].

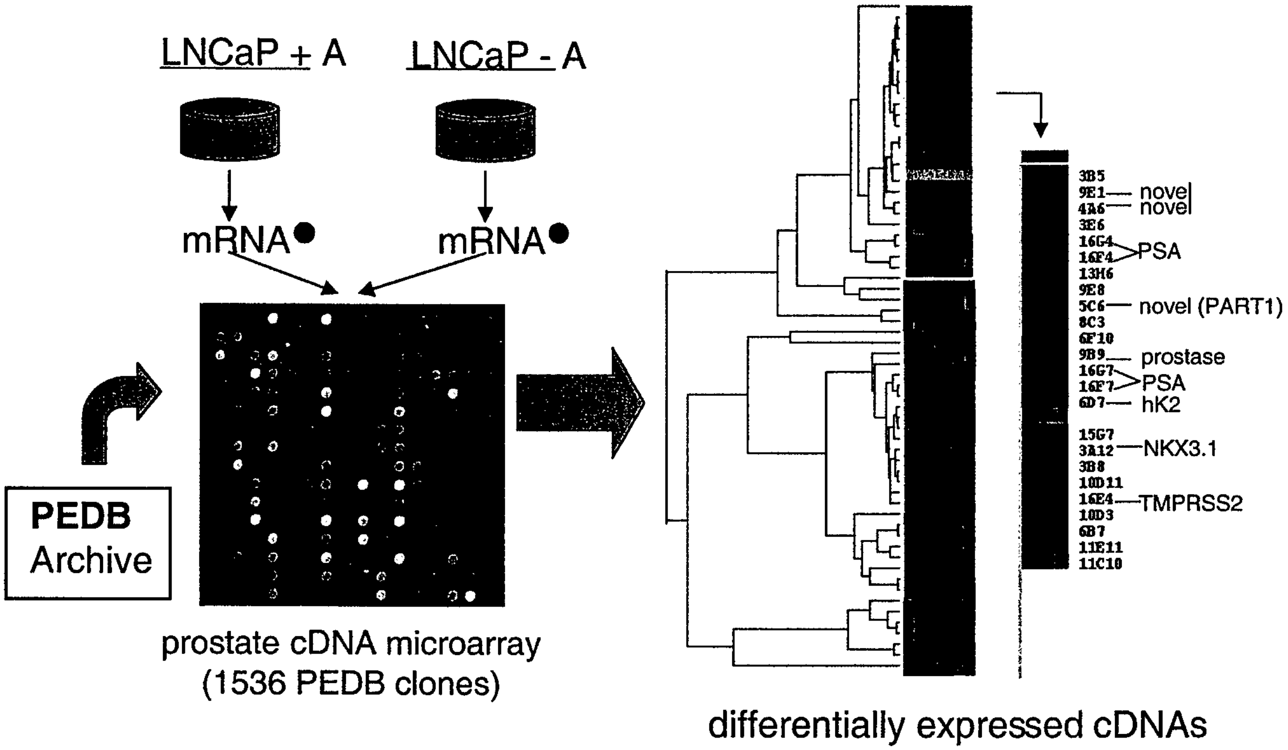


Figure 2. cDNA microarray analysis of prostate androgen-regulated gene expression. A nonredundant clone set comprised of 1536 cDNAs was hybridized with Cy3-labeled (red) cDNA from androgen-stimulated LNCaP cells and Cy5 labeled (green) cDNA from androgen-starved LNCaP cells. The expression ratio for each cDNA was determined and the ratios for all cDNAs with signal intensities 2.33-fold above the standard deviation of the background signal were clustered according to transcript levels over time. The Cluster and TreeView software programs availabe at the Stanford genome web site was used for the analysis (http://rana.Stanford.EDU/software/). Twenty genes were identified with increased expression after androgen stimulation.

3.2 Prostate gene expression analyses: cDNA microarray

Microarrays comprised of 1500 distinct prostate-derived cDNAs were hybridized with fluorescently labeled total cDNA probes produced from androgen-stimulated and androgen-starved LNCaP prostate cancer cells. No cDNAs were identified whose expression level decreased with androgen stimulation. In contrast, the hybridization ratios of 20 different cDNAs were consistently increased by > 2-fold in androgen-stimulated relative to androgen-starved cells (Fig. 2). This group included cDNAs encoding the human glandular kallikrein 2 (hK2) and human glandular kallikrein 3 (hK3), also known as prostate-specific antigen (PSA). The regulation of hK2 and PSA has previously been shown to be mediated by androgens through a mechanism involving androgen-response element (ARE) binding sites in the promoter regions of these genes [20, 21].

In addition to hK2 and PSA, we identified several other genes previously shown to be androgen-regulated, including the prostate homeobox gene NKX3.1 [22], the serine protease prostase/PRSS17 [23], and two genes involved in lipid metabolism. The microarray analysis also indicated that the expression of the membrane-bound serine protease TMPRSS2 [24] was regulated by androgen. We subsequently confirmed the androgen regulation of TMPRSS2 by Northern analysis, identified a putative ARE in the TMPRSS2 promoter region, and demonstrated that TMPRSS2 is highly expressed in the prostate gland relative to other human tissues [25]. Several cDNAs corresponding to uncharacterized genes also exhibited transcriptional regulation by androgen (Fig. 2). We have cloned the full-length cDNA and confirmed the androgen regulation of one of these novel sequences and designated it as PART-1, for Prostate Androgen-Regulated Transcript-1, as it lacks significant homology to nucleotide or protein sequences in the nonredundant subdivision of the GenBank and SWISS-Prot databases [26]. Interestingly, the tissue pattern of PART-1 expression is also essentially restricted to the prostate. The cloning and characterization of the other identified androgen-regulated cDNAs is in progress.

We anticipate that expanding these studies to include a greater portion of the prostate transcriptome coupled with experiments designed to determine direct *versus* indirect transcriptional regulation, and ultimately translational and post-translational regulation of these genes, will establish a framework for understanding the cellular functions mediated by androgens. Despite the important influence of androgenic hormones on prostate cancer growth, relatively few downstream targets of the AR pathway have been described. Studies designed to identify genes regu-

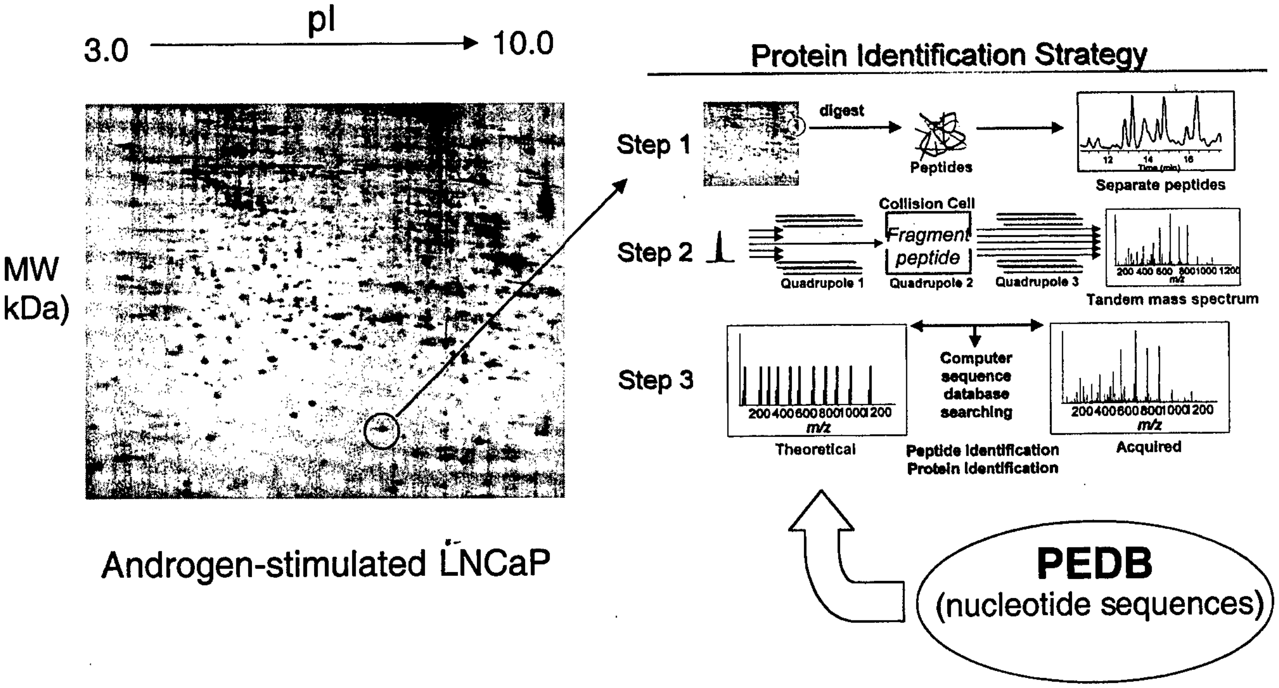
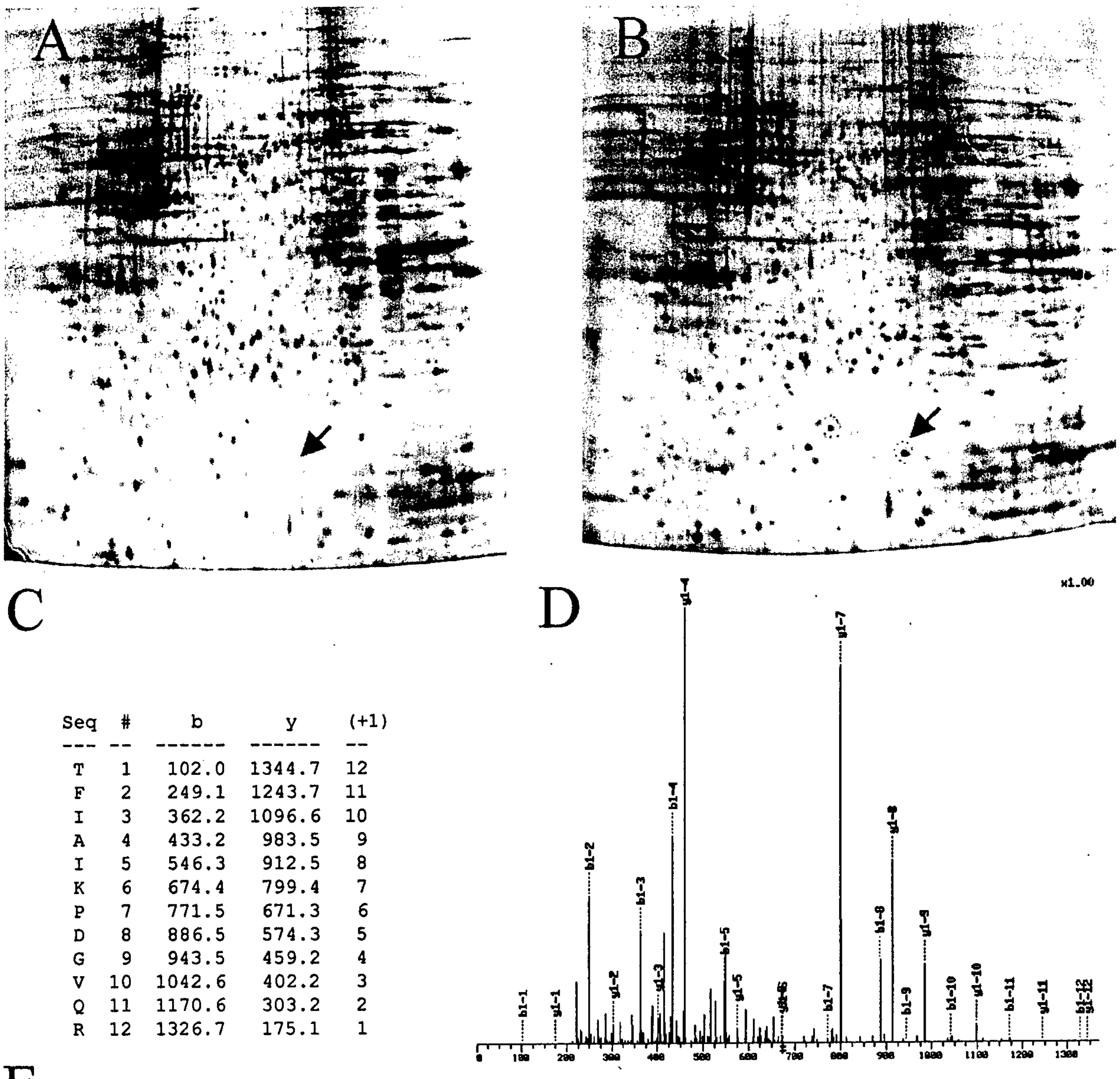


Figure 3. (Left) LNCaP 2-DE protein expression profile with androgen stimulation. (Right) Three-step schema for protein identification using MS and computer sequence database searching.



Seq	#	b	y	(+1)
T	1	102.0	1344.7	12
F	2	249.1	1243.7	11
I	3	362.2	1096.6	10
A	4	433.2	983.5	9
I	5	546.3	912.5	8
K	6	674.4	799.4	7
P	7	771.5	671.3	6
D	8	886.5	574.3	5
G	9	943.5	459.2	4
V	10	1042.6	402.2	3
Q	11	1170.6	303.2	2
R	12	1326.7	175.1	1

E

#	Rank/Sp	(M+H)+	Cn	deltCn	C*10^4	Ions	Reference	Peptide
1.	1 / 1	1344.8	1.0000	0.0000	2.9296.	14/22	NDKA_HUMAN	(R)TFIAIKPDGVQR
2.	2 / 30	1344.6	0.4673	0.5327	1.3689	9/20	RELB_HUMAN	(R)GMPFRYECEGR
3.	3 / 6	1344.8	0.4663	0.5337	1.3661	9/18	VE2_HPV25	(K)RRSRSRSRRK
4.	4 / 12	1344.8	0.4590	0.5410	1.3448	9/24	HIV1U16769	(R)GTKALTEVIALTK
5.	5 / 69	1345.7	0.4519	0.5481	1.3240	7/24	RU2A_HUMAN	(K)TFNPGAGLPTDKK
6.	6 / 8	1345.7	0.4189	0.5811	1.2273	9/22	HUMPRA	(K)GESGQSWPRLTK

Figure 4. Identification of an androgen-regulated protein from metastatic prostate cancer cells by 2-DE and MS. M12AR cells were (A) starved or (B) stimulated for 24 h with the synthetic androgen R1881 and total cell lysates (40 μg each) were subjected to 2-DE. Protein expression profiles were compared and proteins demonstrating a qualitative expression level differences were subjected to in-gel trypsin digestion, and identified by μLC-MS/MS analysis. (C), (D), MS/MS spectrum of identified peptide, peptide sequence, and identified ion series. (E) Results from correlation of acquired peptide fragmentation spectra with database entries (using SEQUEST software). The MS/MS spectrum in (D) was identified as NDKA_HUMAN (nm23) taken from the selected 2-D gel spot. Two additional peptides were identified from this protein in a single run.

lated by androgens in the rat prostate determined that androgens increase the transcription of about 56 genes and decrease the transcription of less than 10 genes [27]. From a therapeutic standpoint, it would be extremely useful to distinguish and subsequently modulate the relevant molecules in the AR program that mediate the divergent processes of cellular proliferation, cellular differentiation, and apoptosis.

3.3 Prostate gene expression analyses: 2-DE and MS

To complement our prostate transcriptional data and provide a more complete picture of prostate gene expression, we have undertaken a comprehensive analysis of that portion of the prostate proteome regulated by androgenic hormones. Reference protein expression profiles were produced for the LNCaP and M12AR prostate cancer cell lines using 2-DE protein separation techniques under steady-state conditions (Fig. 3). Protein expression profiles from cell lysates under conditions of androgen stimulation and androgen starvation have also been generated. A comparison of 2-DE protein profiles under these various conditions yielded a proteomic signature characterized by a subset of proteins with qualitative and quantitative changes. Individual proteins were identified using a sequential process of in-gel trypsin digestion and extraction, peptide separation by μLC, generation of MS/MS spectra, and database correlation with the acquired peptide fragmentation pattern (Fig. 3).

A comprehensive analysis of androgen-induced proteomic signatures is ongoing and our initial experiments demonstrate the utility of this approach in identifying molecules of potential importance in understanding androgen-mediated regulation of prostate cancer progression and metastasis. Figure 4 depicts a portion of the 2-DE protein profile from androgen-starved and androgen-stimulated M12AR prostate cancer cells with a differentially expressed protein spot that is upregulated in M12AR cells after exposure to androgens. This protein was identified as human nucleoside diphosphate kinase A (NDKA/nm23), a well-characterized gene with tumor metastasis suppressor activity in several different human tumors including melanoma, breast, ovary and prostate [28, 29]. Transfection of the DU-145 prostate cancer cell line with NDKA/nm23 inhibited the adhesion to cell matrix and impaired colony growth in soft agar [29].

The M12 prostate cancer cell line is highly tumorigenic when implanted into nude mice and metastasizes to different anatomical sites. Transfection of these cells with a functional androgen receptor (M12AR) markedly decreases the proliferation rate, tumor growth, invasiveness, and *in vivo* metastatic potential when these cells are injected into the prostate glands of nude mice (S. Plymate, unpublished observation). NDKA/nm23 transcripts have been shown to increase rapidly in prostate cancer cell lines after the administration of androgens, though no functional ramifications of this increased expression were described [30].

A possible mechanism for the decreased tumorigenic and metastatic capability of M12AR cells compared with M12 cells lacking the AR involves the upregulation of NDKA/nm23 by androgens through a functional androgen-response program restored by the AR transfection and expression. Such an observation has direct clinical relevance. Both human and *in vitro* studies suggest that there may be a survival benefit from maintaining an androgen responsive cohort of prostate tumor cells [31–33]. This concept has been studied in the LNCaP cell system by comparing the rate of tumor growth in castrated mice implanted with LNCaP cells with subsequent tumor growth (i) without further therapy, or (ii) followed by intermittent androgen replacement. The rate of tumor growth as measured by serum PSA was slower in animals treated with intermittent androgen supplementation compared to those maintained in the castrated state [31].

4 Concluding remarks

The results presented here demonstrate the utility of global expression studies to simultaneously identify multiple genes and gene products of biological relevance that participate in specific metabolic pathways. Both known and unknown genes are rapidly identified. Notable advantages of the microarray-based transcript profiling approach include the ability to perform detailed time-course or variable drug-dose experiments in a robust economical fashion. Controlled replicate experiments can determine system and procedural errors. However, this approach is absolutely dependent upon the identification of diverse clone sets for array construction that are biologically relevant to the system under study. In addition, a significant limitation of transcript profiling methods is the lack of a tight correlation between gene activity as measured by mRNA level, and protein abundance [34]. Global protein analyses focus on the actual biological effector molecules, but are restricted by difficulties in detecting low abundance proteins, accurately measuring the differences in protein levels between two samples, and a dependency on comprehensive annotated sequence databases for protein identification.

Integrating the assembly and annotation of sequence databases with transcript profiling and proteome analyses combines complementary robust approaches that capitalize on the strengths and avoid the limitations of relying on one method. The further expansion of this work to include the analysis of the entire prostate transcriptome coupled with quantitative proteome studies should enable the characterization of gene networks and cellular pathways that can be exploited for therapeutic intervention.

This work was supported in part by the CaPCURE Foundation, and a grant (CA75173) from the National Cancer Institute, grants (DAMD 17-98-1-8499 and PC991274) from the Department of Defense to PSN, a grant (R01-DK52683) and Veterans Affairs Merit Review Program to SRP, the Science Technology Center for Molecular Biotechnology, and a grant (NIH CRR11823) to RA. We wish to thank Joy Ware for the M12 prostate cancer cell line, Steve Lasky and the UW Molecular Biotechnology sequencing center for DNA sequencing support, and Roger Bumgarner, Nigel Clegg, Victoria Hawkins, and Burak Eroglu for bioinformatics support.

Received December 8, 1999

5 References

[1] Denmeade, S. R., Lin, X. S., Isaacs, J. T., *Prostate* 1996, *28*, 251–265.

[2] Isaacs, J. T., Wake, N., Coffey, D. S., Sandberg, A. A., *Cancer Res.* 1982, *42*, 2353–2371.

[3] Bruchovsky, N., Brown, E. M., Coppin, C. M., Goldenberg, S. L., Le Riche, J. C., Murray, N. C., Rennie, P. S., *Progr. Clin. Biol. Res.* 1987, *239*, 347–387.

[4] Velculescu, V. E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M. A., Bassett, D. E., Hieter, P., Vogelstein, B., Kinzler, K. W., *Cell* 1997, *88*, 243–251.

[5] Hawkins, V., Doll, D., Bumgarner, R., Smith, T., Abajian, C., Hood, L., Nelson, P. S., *Nucleic Acids Res.* 1999, *27*, 204–208.

[6] Nelson, P. S., Ng, W. L., Schummer, M., True, L. D., Liu, A. Y., Bumgarner, R. E., Ferguson, C., Dimak, A., Hood, L., *Genomics* 1998, *47*, 12–25.

[7] Strausberg, R. L., Dahl, C. A., Klausner, R. D., *Nature Genet.* 1997, *15*, 415–416.

[8] Huang, X., *Genomics* 1996, *33*, 21–31.

[9] Webber, M. M., Bello, D., Quader, S., *Prostate* 1997, *30*, 58–64.

[10] Bae, V. L., Jackson-Cook, C. K., Maygarden, S. J., Plymate, S. R., Chen, J., Ware, J. L., *Prostate* 1998, *34*, 275–282.

[11] Bae, V. L., Jackson-Cook, C. K., Brothman, A. R., Maygarden, S. J., Ware, J. L., *Int. J. Cancer* 1994, *58*, 721–729.

[12] Garrels, J. I., Franza Jr., B. R., *J. Biol. Chem.* 1989, *264*, 5283–5298.

[13] Blum, H., Beier, H., Gross, H., *Electrophoresis* 1987, *8*, 93–99.

[14] Corthals, LG., Gygi, S. P., Aebersold, R., Patterson, S. D., in: Rabilloud, T. (Ed.), *Proteome Research: 2D Gel Electrophoresis and Detection Methods*, Springer, New York 1999, pp. 197–232.

[15] Yates III, J. R., Eng, J. K., McCormack, A. L., Schieltz, D., *Anal. Chem.* 1995, *67*, 1425–1436.

[16] Hastie, N. D., Bishop, J. O., *Cell* 1976, *9*, 761–774.

[17] Bishop, J. O., Morton, J. G., Rosbash, M., Richardson, M., *Nature* 1974, *250*, 199–204.

[18] Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., Vogelstein, B., Kinzler, K. W., *Science* 1997, *276*, 1268–1272.

[19] Nelson, P. S., Hawkins, V., Schummer, M., Bumgarner, R., Ng, W. L., Ideker, T., Ferguson, C., Hood, L., *Genet. Anal.* 1999, *15*, 209–215.

[20] Riegman, P. H., Vlietstra, R. J., van der Korput, J. A., Brinkmann, A. O., Trapman, J., *Mol. Endocrinol.* 1991, *5*, 1921–1930.

[21] Murtha, P., Tindall, D. J., Young, C. Y., *Biochemistry* 1993, *32*, 6459–6464.

[22] He, W. W., Sciavolino, P. J., Wing, J., Augustus, M., Hudson, P., Meissner, P. S., Curtis, R. T., Shell, B. K., Bostwick, D. G., Tindall, D. J., Gelmann, E. P., Abate-Shen, C., Carter, K. C., *Genomics* 1997, *43*, 69–77.

[23] Nelson, P. S., Gan, L., Ferguson, C., Moss, P., Gelinas, R., Hood, L., Wang, K., *Proc. Natl. Acad. Sci. USA* 1999, *96*, 3114–3119.

[24] Paoloni-Giacobino, A., Chen, H., Peitsch, M. C., Rossier, C., Antonarakis, S. E., *Genomics* 1997, *44*, 309–320.

[25] Lin, B., Ferguson, C., White, J. T., Wang, S., Vessella, R., True, L. D., Hood, L., Nelson, P. S., *Cancer Res.* 1999, *59*, 4180–4184.

[26] Lin, B., White, J. T., Ferguson, C., Bumgarner, R., Friedman, C., Trask, B., Ellis, W., Lange, P., Hood, L., Nelson, P. S., *Cancer Res.*, in press.

[27] Wang, Z., Tufts, R., Haleem, R., Cai, X., *Proc. Natl. Acad. Sci. USA* 1997, *94*, 1299–3004.

[28] Freije, J. M., MacDonald, N. J., Steeg, P. S., *Biochem. Soc. Symp.* 1998, *63*, 261–271.

[29] Lim, S., Lee, H. Y., Lee, H., *Cancer Lett.* 1998, *133*, 143–149.

[30] Yoshimura, I., Wu, J. M., Chen, Y., Ng, C., Mallouh, C., Backer, J. M., Mendola, C. E., Tazaki, H., *Biochem. Biophys. Res. Commun. 1995, 208, 603–609.*

[31] Sato, N., Gleave, M. E., Bruchovsky, N., Rennie, P. S., Goldenberg, S. L., Lange, P. H., Sullivan, L. D., *J. Steroid Biochem. Mol. Biol.* 1996, *58*, 139–146.

[32] Grossfeld, G. D., Small, E. J., Carroll, P. R., *Urology* 1998, *51*, 137–144.

[33] Oliver, R. T., Williams, G., Paris, A. M., Blandy, J. P., *Urology* 1997, *49*, 79–82.

SEQUENCE DATABASES AND MICROARRAYS AS TOOLS FOR IDENTIFYING PROSTATE CANCER BIOMARKERS

LYNETTE H. GROUSE, PETER J. MUNSON, AND PETER S. NELSON

ABSTRACT

Identification, acquisition, and assessment of molecular markers that could be adopted as surrogate endpoints for evaluating a response to prostate cancer intervention strategies is highly desirable. Recent advances in the fields of genomics and biotechnology have dramatically increased the quantity and accessibility of molecular information that is relevant to the study of prostate carcinogenesis. One major advance involves the construction of comprehensive databases that archive gene sequences and gene expression data. This information is in a format suitable for virtual queries designed to distinguish the molecular differences between normal and cancer cells. A second major advance uses robotic tools to construct microarrays comprising thousands of distinct genes expressed in prostate tissues. Such arrays offer a powerful approach for monitoring the expression of thousands of genes simultaneously and provide access for techniques designed to assess patterns or "fingerprints" of gene expression that may ultimately be used as signatures of response to therapeutic intervention. UROLOGY **57** (Suppl 4A): 154–159, 2001. © 2001, Elsevier Science Inc.

The human genome is estimated to comprise approximately 30,000 to 100,000 genes. To confer developmental and functional specificity, only a fraction of this total is active in a given cell type at a given time, and these expressed genes essentially define the state of that cell. The molecular profile of normal and cancer cells, ie, their set of expressed genes, differs in both qualitative (alternative forms of a gene) and quantitative fashions. Measurement of this profile may predict the phenotypic behavior of such cells more accurately than traditional histologic approaches.

To identify informative biomarkers and suitable intermediate endpoints of disease, it would be advantageous to have a catalog or index of all genes and their cognate proteins that are expressed in normal and neoplastic prostate tissues. This resource could then be rapidly exploited to identify candidate biomarkers for evaluation based on homology to known genes of importance in prostate cancer, gene polymorphisms and mutations, or alterations in gene expression. This review will focus particularly on the use of tissue-specific expressed sequence tag (EST) databases, the development and use of cDNA microarrays, and statistical issues related to microarray analyses. These approaches may become essential for identifying new biomarkers in prostate cancer.

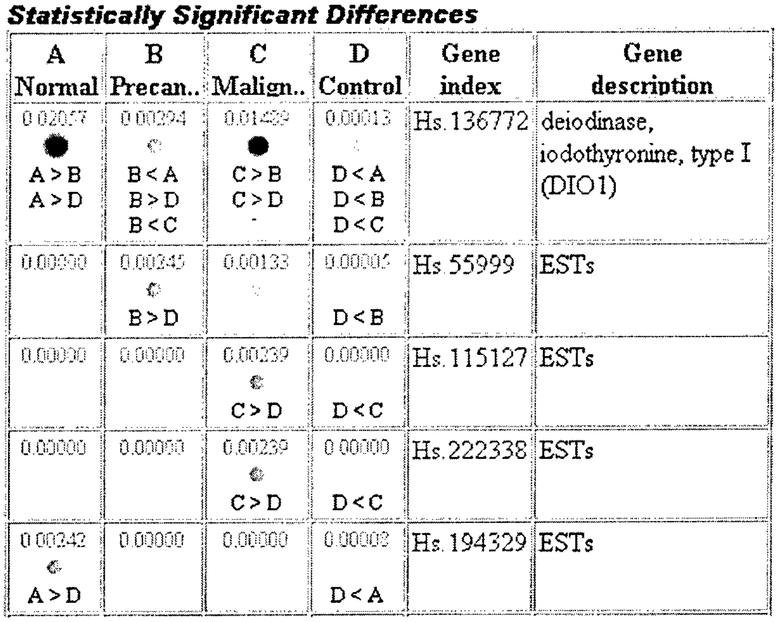
DATABASES AS TOOLS FOR BIOMARKER IDENTIFICATION

In 1997, the National Cancer Institute announced a bold new initiative, the Cancer Genome Anatomy Project (CGAP), with the overall goal of achieving the comprehensive molecular characterization of normal, precancerous, and cancerous cells.[1–3] The CGAP is an interdisciplinary program that uses National Institutes of Health intramural research teams, academic centers, and commercial resources to establish an index of genes expressed in tumors. The CGAP serves as an interface between genomics and cancer research. The new technologies supported by this initiative, and the products resulting from these technologies, will be accessible to the public through an Internet website (http://www.cgap.nci.nih.gov). This Internet site provides information about cDNA libraries of

From the Office of Cancer Genomics, National Cancer Institute (LHG), and the Mathematical and Statistical Computing Laboratory, Center for Information Technology (PJM), National Institutes of Health, Bethesda, Maryland, USA; and the Division of Human Biology, Program in Genomics, Fred Hutchinson Cancer Research Center (PSN), Seattle, Washington, USA

Reprint requests: Lynette H. Grouse, PhD, Office of Cancer Genomics, National Cancer Institute, National Institutes of Health, 31 Center Drive, MSC 2590, Building 31, Room 11A03, Bethesda, MD 20892

0090-4295/01/$20.00
PII S0090-4295(00)00963-8



Statistically Significant Differences

	A Normal	B Precan..	C Malign..	D Control	Gene index	Gene description
1	0.02057 ● A > B A > D	0.00294 • B < A B > D B < C	0.01429 ● C > B C > D	0.00013 · D < A D < B D < C	Hs.136772	deiodinase, iodothyronine, type I (DIO1)
2	0.00000	0.00245 • B > D	0.00133 ·	0.00005 D < B	Hs.55999	ESTs
3	0.00000	0.00000	0.00239 • C > D	0.00000 D < C	Hs.115127	ESTs
4	0.00000	0.00000	0.00239 • C > D	0.00000 D < C	Hs.222338	ESTs
5	0.00242 • A > D	0.00000	0.00000	0.00002 D < A	Hs.194329	ESTs

FIGURE 1. *Gene expression profiles in normal, precancerous, and malignant prostate tissue. Differences in gene expression between cDNA libraries prepared from various types of prostate tissues can be analyzed by using the Digital Differential Display (DDD) software program. Library A is prepared from normal prostate epithelium, Library B from precancerous prostate tissue, Library C from malignant prostate cancer, and Library D is from a control library prepared from a pool of brain, liver, and spleen tissue. The Gene Index contains the UniGene Cluster Identifier, and Gene Description lists the gene name. In each box, the number at the top represents the fraction of sequences in that cDNA library that expresses the gene or EST. The dot is a visual aid, which reflects the numerical values. Each library is compared with each of the other libraries in pairwise analysis. If the difference in gene expression between two libraries is statistically significant, it is indicated by a greater than or less than symbol.*

normal and cancerous tissue, description of the methods used in preparing each library, and informatics tools to perform analyses of gene expression using cDNA library data.

A goal of the CGAP is to facilitate the identification of possible molecular biomarkers for various types of cancer. To enable investigators to analyze molecular databases that are very large and complex, CGAP has developed software tools in collaboration with the National Center for Biotechnology Information at the National Institutes of Health. These software tools aid in the analysis and comparison of gene expression in a variety of tissues and stages of cancer. All of these tools are available on the CGAP Internet website.

An example of a software analysis tool is Digital Differential Display (DDD).[4] DDD is used to compare sequence-based gene expression profiles among individual cDNA libraries or pools of libraries from the same or different tissues. Analysis of different gene expression profiles may identify genes that contribute to a cell's unique characteristics. Such genes, when expressed at different levels in normal and cancer cells, may be considered as candidate biomarkers for use in cancer screening. DDD uses a statistical comparison of genes expressed in each cDNA library to determine which differences are statistically significant. The statistical analysis is based on the Fisher exact test.[5] Differences in gene expression values are presented both visually and numerically.

An example of a DDD analysis of three cDNA libraries made from prostate tissue is shown in Figure 1. Row 1 shows an expression profile of a gene that has a known function, whereas rows 2 to 5 show expression differences between genes of unknown function, referred to as ESTs. Row 1, column D, shows that all three prostate cDNA libraries have increased expression of the DIO1 gene compared with that of control. Within the prostate libraries, column A shows increased gene expression compared with that of the precancerous library in column B, but was not shown to be statistically significantly increased when compared with malignant prostate cancer tissue libraries. The power of this analysis is in identification of possible biomarkers within anonymous EST sequences. In row 2, the expression of this EST is increased over control in only the precancerous prostate cDNA library, whereas the EST in rows 3 and 4 is increased only in cancerous prostate tissues. These genes could be evaluated as candidate biomarkers to identify prostate cancer disease progression. The Prostate Expression Database (PEDB) (http://www.mbt.washington.edu/PEDB)[6] is another online resource of prostate genetic information. The PEDB is a curated relational database and suite of analysis tools designed specifically for the study of prostate gene expression in normal and diseased states. The ESTs, derived from more than 40 human prostate cDNA libraries, are assembled into distinct species groups that are annotated with information from the GenBank, dbEST, and Unigene public sequence databases. The expression pattern of each gene can be viewed across all libraries or tissues using the Virtual Expression Analysis Tool (VEAT), a graphical user interface written in Java for intra- and interlibrary gene expression analyses.

cDNA EXPRESSION ARRAYS FOR BIOMARKER IDENTIFICATION

The inherent heterogeneity of prostate cancers and the diversity of therapeutic interventions suggest that it is unlikely that a single biomarker or intermediate endpoint that will provide sufficient sensitivity or specificity for assessing a treatment response can be identified. Efforts have been directed toward methods of simultaneously measuring multiple biomarkers at the DNA, RNA, or protein levels. Such a multiplexed approach will greatly expand the information gained from each patient sample and clinical trial. In addition, patterns in biomarker data may be identified that together exceed the sum of individual measurements.

Recent developments in informatics, miniaturization, and robotics have provided new extremely powerful approaches for comprehensive measurements of genetic alterations that occur in neoplasia. These measured alterations could also reflect a response (or lack of response) to a chemopreventive or therapeutic agent. One such comprehensive approach involves the use of DNA arrays, a technique that combines the proven chemistry of nucleic acid hybridization with advanced automation and imaging technology to quantitatively detect changes in the expression levels of thousands of genes simultaneously. DNA arrays have been assembled in several configurations, including oligonucleotide arrays,[7] microarrays of cDNA spotted on glass slides,[8] and DNA spotted onto nylon membranes.[9] The basic method is straightforward: DNA representing a particular gene of interest is either spotted (printed) or synthesized onto a solid support, such as a glass microscope slide, silicon wafer, or nylon membrane (Figure 2). The procedure is repeated in an automated fashion with thousands of different genes, such that each is deposited in a precise spatial location that allows for the subsequent identification of any individual spot. Probes representing the expressed genetic information in a tissue sample are labeled with radioactive or fluorescent markers that can be quantified by sensitive detectors and used for comparative analyses. A limitation on the number of individual elements that can be placed on the area of a given "chip" array places a premium on efficient construction. This is accomplished by eliminating redundancy (maximizing diversity), and incorporating DNA sequences that are relevant for the biological system under study.

Gene expression catalogs, such as the CGAP and the PEDB,[6,10] can be exploited for the construction and analysis of cDNA expression arrays by providing a virtual archive of thousands of genes expressed in prostate tissue. Coupling this virtual repository with the physical clones representing the corresponding DNA molecules allows for the construction of comprehensive arrays. The continued expansion of this resource to encompass all prostate transcripts will allow for the simultaneous analysis of all genes expressed in normal and neoplastic prostate cells. This effort will require extensive testing on prostate tumors and a further refinement of the methods to include statistical measures of biological and experimental variance.

STATISTICAL ISSUES IN THE ANALYSIS OF cDNA MICROARRAYS

Special-purpose, tissue-specific cDNA microarrays can now be routinely generated using commercially available spotting robots, either using glass-based or nylon-based substrate. A growing number of commercial cDNA microarrays are also available, giving smaller labs the opportunity to use this technology. Careful attention to the design and statistical analysis of each experiment is essential, especially given the high cost of microarrays. As with any assay procedure, microarray data are subject to three major sources of random and systematic error: reagent quality, sample preparation, and laboratory technique. The most important reagent is the microarray itself, which may be subject to significant batch-to-batch variability. The array may include clones of questionable quality, possibly including troublesome repetitive DNA or another contaminating sequence. Variability of the substrate, either nylon or glass, can have a marked effect on the uniformity of the array image. Sample preparation includes all tissue handling, cell isolation, RNA extraction, and labeling steps. Unintended variation in RNA content may easily result from poor temperature control, heat shock, degra-

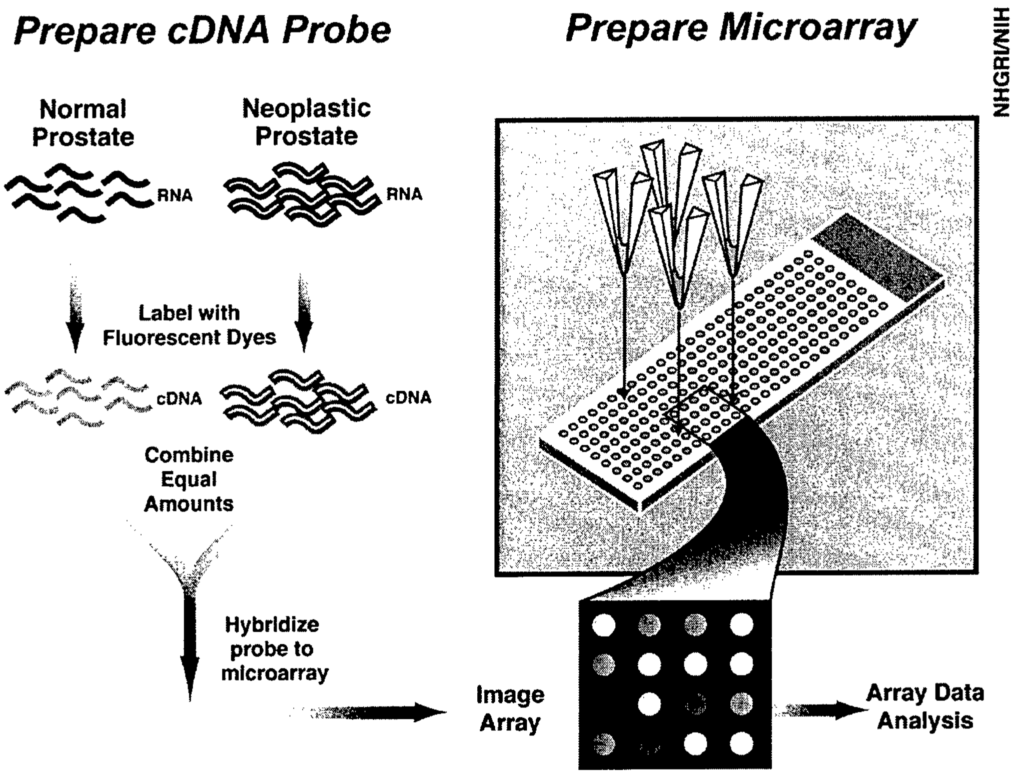


FIGURE 2. *cDNA microarray construction and analysis. Microarray assays are performed in a multistep process. First, microarrays are prepared by assembling sets of cDNA clones in 96- or 384-well microtiter plates. Small tweezer tips or needles attached to a robotic arm are used to withdraw small amounts of the DNA solution from the microtiter plates and print them onto glass microscope slides in a precise spatial orientation with high replicative fidelity. cDNA probes are prepared from two distinct tissue sources (eg, normal tissue and neoplastic tissue) by first extracting RNA followed by a conversion step to cDNA that incorporates a different fluorescent dye into the different tissue source cDNA (eg, green for normal and red for neoplastic). These labeled cDNA probes are then combined and hybridized to the microarray such that cDNAs in the probe will attach to their complementary cDNA spot on the microarray surface. Nonhybridizing cDNAs are removed by a washing step, and the remaining bound cDNA molecules are quantitated by measuring the fluorescent intensity at every spot location. Array analyses determine the ratio of intensities at each spot and thus identify specific genes that are overexpressed in normal tissue relative to neoplastic (green spot), overexpressed in neoplastic relative to normal (red spot), or expressed at equivalent levels (yellow spot).*

dation, sample handling, etc. Fluorescence or radioactive label incorporation may also be subject to variation and can strongly influence the results. During the hybridization of labeled probe to the target cDNA on the array, carefully controlled time, temperature, and agitation conditions should prevail. Issues of saturation and dynamic range compression may arise during image acquisition and storage.

By far the most straightforward way to address each of these issues is by use of independent replicated experiments. Apparent gene expression changes that persist through such repeated experiments can correctly be ascribed to interesting biological changes rather than artifacts of the assay itself. The following illustrative analysis of duplicate experiments easily screens out many artifactual expression changes. We compared a melanoma cell line to a prostate tumor cell line for expression differences on a prostate-specific, nylon-based cDNA array.[11] Spot intensities were quantified using the P-SCAN software (available at http://abs.cit.nih.gov/PSCAN). The intensities of each spot were compared in Figure 3A, which at first seems to indicate that a large number of genes have greater than fourfold changes in relative expression levels between the two cell types. Analysis of a duplicate experiment gave a different picture. Figure 3B shows that a much smaller number of genes undergo greater than fourfold changes consistently in both experiments, meaning that many of the apparent fourfold changes in the first experiment were "false-positives." A family of differen-

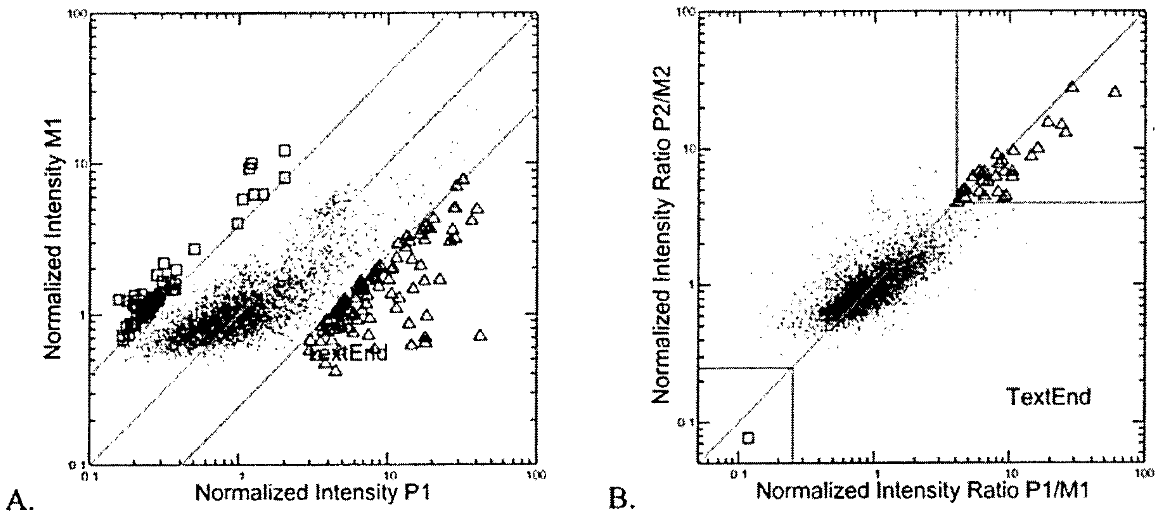


FIGURE 3. *Comparison of melanoma (M) expression levels to that from a prostate tumor cell line (P).*[11] *(A) Normalized intensities show more than 148 genes with apparent expression change over four-fold (up, squares or down, triangles). (B) Expression ratios are compared for duplicate experiments (P1/M1, P2/M2). Only 31 genes are consistently over- or underexpressed by greater than fourfold in both experiments.*

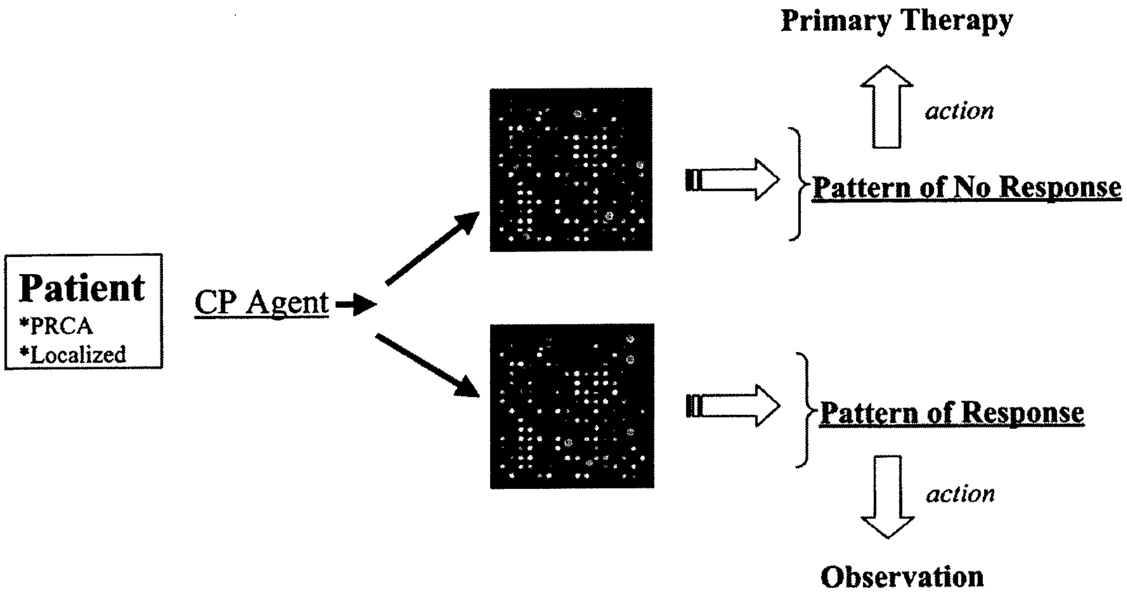


FIGURE 4. *Molecular profiles as surrogate endpoint biomarkers. Patients with localized prostate cancer are treated with a chemopreventive (CP) agent. A biopsy is performed and subjected to molecular profiling by cDNA microarray analysis. The pattern of expression is compared with reference patterns previously shown to correlate with a tumor response or lack of tumor response to the CP agent. These data are used to guide further therapeutic intervention.*

tially expressed genes also clearly emerged and was later confirmed by Northern blot analysis. Reduction in the number of false-positives can be essential when using microarray technology to look for new cancer markers, as tens of thousands of clones must be screened.

APPLICATION TO CHEMOPREVENTION

Among their many applications, database and array-based methods of genetic analysis can be useful for the identification, acquisition, and assessment of candidate molecular markers that could be adopted as surrogate endpoints for assessing preventive strategies (chemoprevention or nutritional intervention). One scenario involves a cohort of patients diagnosed with low- or intermediate-grade prostate cancers by needle biopsy. Patients who elect to forgo primary therapy (radical prostatectomy or radiotherapy) could be offered a chemopreventive agent aimed at halting cancer progression. Gene expression profiles of tumor tissue before and after the chemopreventive agent would be assessed for expression patterns correlating with a propensity for the cancer to progress, indicating that a primary therapy should be offered, or for the cancer to respond to the chemopreventive agent and thus require no further intervention (Figure 4). The development of this type of assay is clearly desirable, but defining predictive patterns of expression is not a trivial task.

ACKNOWLEDGMENT. The authors thank Darryl Leja for his efforts in creating the microarray figure for this manuscript.

REFERENCES

1. Strausberg RL, Dahl CA, and Klausner RD: New opportunities for uncovering the molecular basis of cancer. Nat Genet **15**: 415–416, 1997.

2. Strausberg RL: Genetics in profile. Trends Genet **14**: 50–51, 1998.

3. Strausberg RL: The Cancer Genome Anatomy Project: building a new information and technology platform for cancer research, in Srivastava S (Ed): *Molecular Pathology of Early Cancer*. Amsterdam, IOS Press, 1999, pp 365–370.

4. Spouge JL: Digital Differential Display. Available at: http://www.ncbi.nlm.nih.gov/ncicgap/cgapdeliv.cgi?title=CGAP+Digital+Differential+Display&ins_file=ddd.html_frag. Source: Cancer Genome Anatomy Project (CGAP) website. Accessed: October 2000

5. Spouge JL: Fisher exact test. Available at: http://www.ncbi.nlm.nih.gov/ncicgap/cgapdeliv.cgi?title=Fisher+Exact+Test&ins_file=fisher.html_frag. Source: Cancer Genome Anatomy Project (CGAP) website. Accessed: October 2000

6. Hawkins V, Doll D, Bumgarner R, *et al*: PEDB: the Prostate Expression Database. Nucleic Acids Res **27**: 204–208, 1999.

7. Chee M, Yang R, Hubbell E, *et al*: Accessing genetic information with high-density DNA arrays. Science **274**: 610–614, 1996.

8. Schena M, Shalon D, Davis RW, *et al*: Quantitative monitoring of gene expression patterns with a complementary DNA microarray [see comments]. Science **270**: 467–470, 1995.

9. Nguyen C, Rocha D, Granjeaud S, *et al*: Differential gene expression in the murine thymus assayed by quantitative hybridization of arrayed cDNA clones. Genomics **29**: 207–216, 1995.

10. Nelson PS, Clegg N, Eroglu B, *et al*: The Prostate Expression Database (PEDB): status and enhancements in 2000. Nucleic Acids Res **28**: 212–213, 2000.

11. Carlisle AJ, Prabhu VV, Elkahloun J, *et al*: Development of a prostate cDNA microarray and statistical gene expression analysis package. Mol Carcinog **28**: 12–22, 2000.